



Formation et développement IPUMS: la demande de données (Exercice 1 pour SAS)



Objectifs

- Créer et télécharger un extrait de données IPUMS PMA
- Décompresser le fichier de données et lire les données dans Stata
- Analyser les données en utilisant un exemple de code

Résumé

Comprendre comment les ensembles de données IPUMS PMA des ménages et des femmes sont structurés et comment les exploiter pour explorer vos intérêts de recherche. Cet exercice utilisera l'ensemble de données IPUMS PMA pour explorer la relation entre l'arrêt du planning familial et la richesse. Vous allez créer un extrait de données avec les variables LINENO, WEALTHQ, FPSTOPWHY, FPNOWUSPILL, FPAGE1STUSE, FP1STMETHOD, FQWEIGHT, POPWT, HQWEIGHT; puis vous allez utiliser le code d'exemple à analyser les données.

Code SAS à réviser

Code	Procédure
<u>PROC FREQ;</u>	Générer une procédure de fréquence
<u>PROC MEANS;</u>	Générer une procédure de moyen, rendre le moyen d'une variable
<u>tables</u>	Syntaxe requise pour afficher les résultats en format tableau
<u>where</u>	Sélectionner certains cas à inclure dans la procédure
<u>weight;</u>	Fournir la pondération de chaque observation dans l'ensemble de données saisies
<u>data newdata; set olddata;</u>	Utiliser le « data step » pour renommer ou classer les données selon opérateurs conditionnels, commandes prémisses-conclusion, ou transformation.
<u>PROC UNIVARIATE;</u>	A utiliser afin de créer histogrammes

Erreurs courantes à éviter

- Pas complètement décompresser les données.
- Ne pas changer le répertoire de travail ou vos données sont stockées
- Oublier de lancer la procédure avec « run »
- Oublier de terminer la commande avec point-virgule « ; »

Enregistrement avec IPUMS

- Allez sur <http://pma.ipums.org>, cliquez sur Enregistrer pour utiliser IPUMS PMA sur le côté gauche de l'écran. Cliquez sur le bouton S'inscrire à IPUMS PMA et remplissez le formulaire pour demander l'accès. Vous devrez attendre que votre compte soit approuvé pour accéder aux données. Une fois que vous avez reçu l'email d'acceptation, cliquez sur "Connexion" en haut de la page et utilisez votre email et votre mot de passe.



- Aller à « Browse and Select Data »
- Choisissez l'unité d'analyse « Person »

CHOOSE THE UNIT OF ANALYSIS FOR DATA BROWSING	
PERSON	EACH RECORD WILL BE A PERSON DESCRIPTION
SERVICE DELIVERY POINT	EACH RECORD WILL BE A SERVICE DELIVERY POINT DESCRIPTION

- Cliquez sur la case « Select Samples, » cochez la case pour Kenya 2016 R5.

Kenya [2016 R5](#) [2015b R4](#) [2014b R2](#)
 [2015a R3](#) [2014a R1](#)

- Faites défiler vers le bas de la page et cliquez sur l'option du bouton radio pour « All Cases (Respondents and Non-respondents to Household and Female Questionnaires) ». La valeur par défaut est Répondantes femmes.
- Cliquez sur la case « Submit Sample Selections. »

Sample Members

- Female Respondents
- Female Respondents and Household Members
- Female Respondents and Female Non-respondents
- All Cases (Respondents and Non-respondents to Household and Female Questionnaires)

SUBMIT SAMPLE SELECTIONS

- À l'aide du menu déroulant ou de la fonction de recherche, sélectionnez les variables suivantes:

LINENO: numéro d'une personne dans l'unité d'échantillonnage

WEALTHQ: quintile de richesse

FPSTOPWHY: raison pour l'arrêt de la méthode de planification familiale

FPAGE1STUSE: l'âge de la première utilisation de planification familiale

FP1STMETHOD: première méthode de planification familiale utilisée



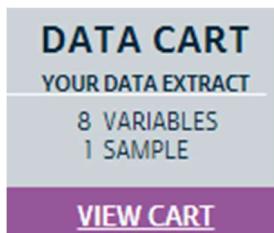
FQWEIGHT: variable de pondération au niveau de la femme

POPWT: facteur d'expansion de population

HQWEIGHT: variable de pondération au niveau du ménage

FPNOWUSPILL: utilise la pilule actuellement

- Cliquez sur le bouton violet VIEW CART sous votre panier de données



- Vérifier la sélection des variables. Notez que certaines variables apparaissent dans votre panier même si vous ne les avez sélectionnés pas. Ils ne sont pas inclus dans le compte de variables dans votre panier de données. Ces variables présélectionnées sont nécessaires pour la pondération, l'estimation de la variance, ou pour identifier l'année, le pays, et le round d'un échantillon.
- Cliquez sur le bouton « Create Data Extract ».
- Examinez l'écran « Extract Request », décrivez votre extrait et cliquez sur « Submit Abstract ».
- Vous recevrez un email lorsque les données seront disponibles pour le téléchargement.
- Pour accéder à la page pour télécharger les données, suivez le lien dans l'e-mail ou le lien « My Data Extracts » sur la page d'accueil.

Lancer les données dans votre logiciel de statistiques

Télécharger les données

Allez sur <http://pma.ipums.org> et cliquez sur «My Data».

Extract		Formatted		Fixed-width Text Files			
Number	Date	Data		Data	Command Files 		
51	2018-10-26	--		Download DAT	SPS	SAS	STATA



- Faites un clic droit sur le lien de données à côté de l'extrait que vous avez créé.
- Choisissez « Save Target As ... » (ou « Save Link As ... »)
- Enregistrer dans « Documents » (qui devrait apparaître comme localisation par défaut)
- Faites la même opération pour le lien SAS à côté de l'extrait

- Trouver le dossier "Documents" dans le menu de démarrage.
- Faites un clic droit sur le fichier «.dta ».
- Dans la fenêtre qui s'ouvre, appuyez sur le bouton « Extract » pour l'extraire ici.
- Vérifier que le fichier "pma_000..." se trouve dans le dossier "Documents"
- Un logiciel de décompression gratuit est disponible à l'adresse
<http://www.irnis.net/soft/wingzip/>
- Ouvrir le fichier « pma_000##.sas »
- Dans la fenêtre du fichier « pma_000##.sas » changer la première ligne de « libname IPUMS ':' » à « libname IPUMS "C:\Users\Documents...;" » en remplaçant le répertoire correspondant au fichier où vos données sont sauvegardées.
- Après « filename ASCIIIDAT », saisir l'emplacement du fichier complet, en terminant en « pma_000##.dat; »
- Choisir « Submit » sous l'option de menu Run.



Analyser l'échantillon

Partie 1: Fréquences de WEALTHQ pour le questionnaire ménage auprès des ménages

1. Selon l'onglet « Description » du site Web, que mesure la variable WEALTHQ? _____

2. Selon l'onglet « Codes, » quelles étiquettes de valeur et quels codes numériques s'appliquent à chaque quintile de richesse? Incluez des étiquettes et des codes expliquant les valeurs « manquantes ». _____
3. Selon l'onglet « Univers, » qui est dans l'univers de WEALTHQ dans l'échantillon Kenya 2016 Round 5? _____
4. Quel pourcentage de tous les membres du ménage est considéré comme vivant dans un ménage du quintile de revenu le plus bas? _____

```
PROC FREQ;  
tables wealthq;  
run;
```

5. Quel pourcentage de ménages sont considérés dans le quintile de revenu le plus bas? _____

```
PROC FREQ;  
tables wealthq;  
where lineno in(1);  
run;
```



Partie 2: Fréquences de FPSTOPWHY chez les femmes

6. Selon l'onglet « Description » sur le site Web, que mesure la variable FPSTOPWHY?

Vous le trouverez dans le groupe de variables « Discontinuation of Family Planning ».

7. Selon l'onglet « Univers, » qui est dans l'univers de FPSTOPWHY dans l'échantillon

Kenya 2016 - Round 5? _____

8. Créez un tableau de fréquences pour la variable FPSTOPWHY pour l'échantillon

Kenya 2016 Round 5. Quelles sont les deux réponses les plus fréquents? Quelle proportion d'enquêtes fournit ces réponses? _____

```
PROC FREQ;  
tables fpstopwhy;  
run;
```

9. Étant donné l'univers indiqué en 7, seulement une petite fraction des femmes âgées de 15 à 49 ans a fourni des réponses à FPSTOPWHY. Les cas hors univers incluent également les hommes et les femmes en dehors de l'âge de procréer. Supposons que vous vouliez savoir quelle proportion de ceux qui ont fourni une réponse à FPSTOPWHY a choisi l'une des deux principales réponses. Comment est-ce que cette réponse diffère-t-elle de votre réponse pour 8?
-

```
PROC FREQ;  
tables fpstopwhy;  
where fpstopwhy <>96 and fpstopwhy<>95 and fpstopwhy<>99;  
run;
```



10. Parmi les femmes de l'échantillon qui ont cessé d'utiliser une méthode de planification familiale au cours de la dernière année en raison de leur prix, quelle proportion vivait dans des ménages appartenant à l'un des deux quintiles de revenu les plus bas? _____

```
PROC FREQ;  
tables fpstopwhy*wealthq / nocol norow nopercent;  
run;
```

Partie 3: Utilisation des pondérations du ménage, de la femme et de la population (HQWEIGHT, FQWEIGHT et POPWT)

Les données IPUMS PMA nécessitent l'utilisation de pondérations pour s'assurer que chaque échantillon est représentatif de la population à partir de laquelle il a été tiré. Toutefois, comme les variables font alternativement référence à la population des ménages ou à la population des femmes âgées de 15 à 49 ans dans chaque pays, les utilisateurs doivent sélectionner des pondérations correspondant à l'unité d'analyse envisagée.

Par exemple: le tableau de fréquence pour FPSTOPWHY de la partie 2D indique que, parmi les femmes âgées de 15 à 49 ans du Kenya, cinquième cycle de 2016, qui ont répondu à cette question, 30,85% ont cessé d'utiliser les méthodes de planification familiale au cours de l'année dernière parce qu'elles voulaient devenir enceintes. Étant donné que l'unité d'analyse pertinente pour FPSTOPWHY comprend les femmes âgées de 15 à 49 ans (plutôt que les ménages), FQWEIGHT doit être utilisé pour tirer des conclusions sur cette population particulière.



11. Utilisez FQWEIGHT pour estimer la proportion de toutes les femmes Kenyanes âgées de 15 à 49 ans qui ont cessé d'utiliser les méthodes de planification familiale afin de tomber enceintes en 2016. _____

```
PROC FREQ;  
tables fpstopwhy;  
weight fqweight;  
run;
```

Bien que FQWEIGHT puisse être utilisé pour estimer les proportions, il ne devrait pas être utilisé pour générer des comptes de population représentatifs. Au lieu de cela, POPWT combine les informations de FQWEIGHT avec les estimations de population nationales de la Division de la Population des Nations Unies pour créer un coefficient de pondération approprié à cet effet (voir https://pma.ipums.org/pma/population_weights.shtml pour des détails).

12. Utilisez POPWT pour estimer le nombre total de femmes Kenyanes âgées de 15 à 49 ans qui ont cessé d'utiliser les méthodes de planification familiale afin de tomber enceintes en 2016. _____

```
PROC FREQ;  
tables fpstopwhy;  
weight popwt;  
run;
```

En revanche, les variables relatives aux caractéristiques du ménage doivent être pondérées avec HQWEIGHT, qui tient compte de la probabilité de chaque ménage d'être échantillonné et de la non-réponse du ménage. Étant donné que chaque membre du



même ménage partage la même valeur pour HQWEIGHT, les utilisateurs doivent également veiller à ce que chaque ménage ne soit représenté qu'une fois dans leur analyse.

Par exemple: la partie 1 a révélé que 18,85% des ménages de l'échantillon du Kenya, série 5, 2016 appartiennent au quintile de revenu le plus bas. Étant donné que la variable WEALTHQ reflète une caractéristique du ménage, HQWEIGHT devrait être utilisé pour tirer des conclusions sur la population des ménages au Kenya en 2016.

13. Utilisez HQWEIGHT pour estimer la proportion de ménages appartenant au quintile de revenu le plus bas au Kenya en 2016. L'échantillon représente-t-il plus ou moins les ménages du quintile de revenu inférieur par rapport à la population exact des ménages?
-

```
PROC FREQ;  
tables wealthq;  
where lineno in(1);  
weight hqweight;  
run;
```

Les utilisateurs doivent noter que, tout comme FQWEIGHT, HQWEIGHT ne doit être utilisé que pour générer des estimations de proportion pour la population des ménages. Il n'existe actuellement aucun coefficient de pondération permettant d'estimer le nombre de fréquences de la population.

Partie 4: Représentation graphique

14. Examinez un tableau de fréquence pour la variable FPAGE1STUSE, qui reflète l'âge auquel le répondant a commencé à utiliser n'importe quelle méthode de planification familiale. Créez un histogramme de 35 intervalles pour FPAGE1STUSE, en excluant les valeurs qui manquent.



```

PROC FREQ;

tables FPAGE1STUSE;

run;

PROC UNIVARIATE data=IPUMS.pma_00030;
var FPAGE1STUSE;
where FPAGE1STUSE<900;
histogram / endpoints=(10 to 50 by 1.14);
run;

```

15. Considérons maintenant la variable FP1STMETHOD, qui classe la première méthode de planification familiale jamais utilisée par un répondant comme étant « moderne » (valeurs 100-199) ou « traditionnelle » (valeurs 200-299). Créez une nouvelle variable qui divise le type de méthode en méthodes modernes et traditionnelles en fonction des codes et des fréquences présentées sur le site internet de PMA. Recréez l'histogramme ci-dessus deux fois, en vous limitant aux méthodes modernes ou traditionnelles. Spécifiez 35 emplacements pour chacun à l'aide de l'option bin (35).

```

data IPUMS.pma_00030;
set IPUMS.pma_00030;
  if fp1stmethode <200 then method_mod = 1;
  else if fp1stmethode >200 & fp1stmethode <900 then method_mod = 0;
  else method_mod = .;
run;

PROC UNIVARIATE data=IPUMS.pma_00030;
var FPAGE1STUSE;

```



```
where FPAGE1STUSE <50 & method_mod=1;  
histogram / endpoints=(10 to 50 by 1.14);  
run;  
  
PROC UNIVARIATE data=IPUMS.pma_00030;  
var FPAGE1STUSE;  
where FPAGE1STUSE <50 & method_mod=0;  
histogram / endpoints=(10 to 50 by 1.14);  
run;
```



RÉPONSES

Partie 1: Fréquences de WEALTHQ pour les enquêtes auprès des ménages

1. Selon l'onglet « Description » du site Web, qu'est-ce que la variable WEALTHQ mesure? Pour les ménages qui ont fini l'enquête, WEALTHQ fait référence à la richesse relative du ménage dans lequel vit la femme, divisée en quintiles du plus pauvre (le quintile le plus bas) au plus riche (le quintile le plus élevé).
2. Selon l'onglet « Codes, » quelles étiquettes de valeur et quels codes numériques s'appliquent à chaque quintile de richesse? Incluez des étiquettes et des codes expliquant les valeurs « manquantes ».

01 – Lowest quintile (Quintile le plus bas)
02 – Lower quintile (Quintile inférieur)
03 – Middle quintile (Quintile moyen)
04 – Higher quintile (Quintile supérieur)
05 – Highest quintile (Quintile le plus élevé)

96 – Not interviewed (household questionnaire) (Non interrogé (questionnaire ménage))
98 – No response or missing (Pas de réponse ou manquant)

3. Selon l'onglet « Univers, » qui est dans l'univers de WEALTHQ dans l'échantillon Kenya 2016 Round 5? Toutes les personnes.
4. Quel pourcentage de tous les membres du ménage est considéré comme vivant dans un ménage du quintile de revenu le plus bas? 22,68%
5. Quel est le pourcentage de ménages énumérés dans le quintile de revenu inférieur? 19,35%

Partie 2: Fréquences de FPSTOPWHY chez les femmes

6. Selon l'onglet « Description » sur le site internet, que mesure la variable FPSTOPWHY? Vous le trouverez dans le groupe de variables Arrêt du Planning Familial. Pour les femmes qui ont récemment utilisé une méthode de planification familiale pour retarder ou éviter une grossesse, mais ne l'utilisent plus, FPSTOPWHY indique la raison de leur arrêt.



7. Selon l'onglet Univers, qui est dans l'univers de FPSTOPWHY dans l'échantillon Kenya 2016 Round 5? Femmes âgées de 15 à 49 ans qui ont utilisé une méthode de planification familiale pour retarder ou éviter une grossesse au cours des 12 derniers mois, mais qui ne les utilisent pas actuellement.
8. Créez un tableau de fréquences pour la variable FPSTOPWHY pour l'échantillon Kenya 2016 Round 5. Quelles sont les deux réponses les plus courantes? Quelle proportion d'enquêtes fournit ces réponses? NIU (hors de l'univers de réponses) (97,23%); Voulait devenir enceinte (0,46%)
9. Étant donné l'univers indiqué en 7, seulement une petite fraction des femmes âgées de 15 à 49 ans a fourni des réponses à FPSTOPWHY. Les cas hors d'univers incluent également les hommes et les femmes en dehors de l'âge de procréer. Supposons que vous vouliez savoir quelle proportion de ceux qui ont fourni une réponse à FPSTOPWHY a choisi l'une des deux principales réponses. En quoi cette réponse diffère-t-elle de votre réponse pour 8? Voulait devenir enceinte (31,25%) Relations sexuelles peu fréquentes / mari absent (22,92%). Ces proportions ont maintenant un dénominateur plus approprié pour l'analyse.
10. Parmi les femmes de l'échantillon qui ont cessé d'utiliser une méthode de planification familiale au cours de la dernière année en raison des prix, quelle proportion vivait dans un ménage appartenant à l'un des deux quintiles de revenu les plus bas? 2/3, soit 66%

Partie 3: Utiliser le coefficient de pondération du ménage, des femmes et de la population (HQWEIGHT, FQWEIGHT et POPWT)

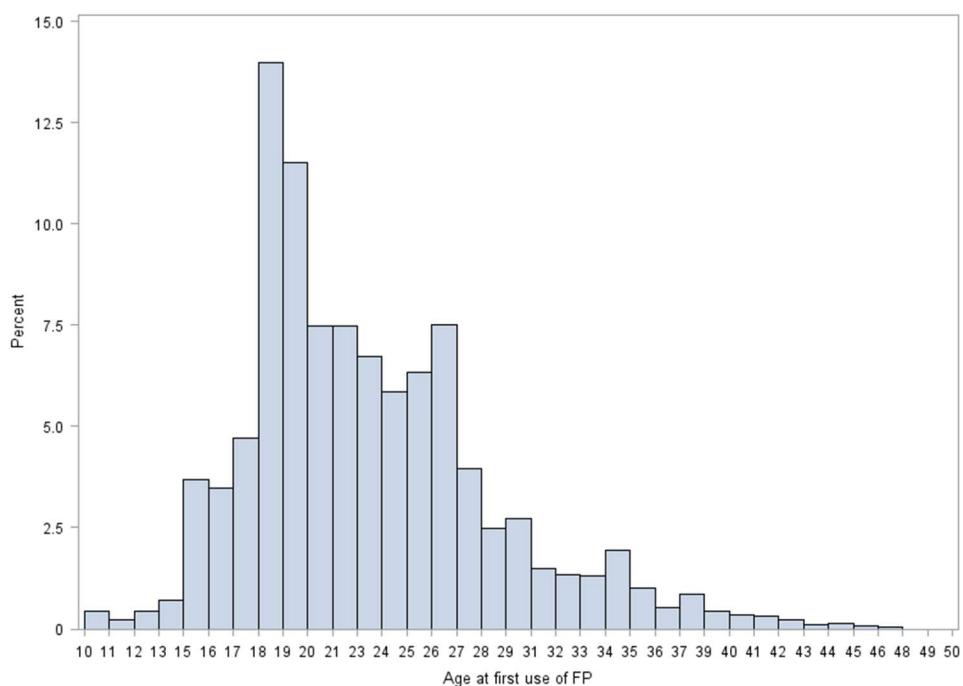
11. Utilisez FQWEIGHT pour estimer la proportion de toutes les femmes Kenyanes âgées de 15 à 49 ans qui ont cessé d'utiliser les méthodes de planification familiale afin de tomber enceintes en 2016. 2.19%



12. Utilisez POPWT pour estimer le nombre total de femmes Kenyanes âgées de 15 à 49 ans qui ont cessé d'utiliser les méthodes de planification familiale afin de devenir enceintes en 2016. 273,016
13. Utilisez HQWEIGHT pour estimer la proportion de ménages appartenant au quintile de revenu le plus bas au Kenya en 2016. L'échantillon représente-t-il plus ou moins les ménages du quintile de revenu inférieur par rapport à la population réelle des ménages? 18,86%. Le quintile de revenu le plus bas est légèrement sur-représenté dans l'échantillon par rapport à la population réelle des ménages.

Partie 4: Représentation graphique

14. Examinez un tableau de fréquence pour la variable FPAGE1STUSE, qui reflète l'âge auquel le répondant a commencé à utiliser une méthode de planification familiale pour la première fois. Créez un histogramme pour FPAGE1STUSE, en excluant les valeurs manquantes.



15. Considérons maintenant la variable FP1STMETHOD, qui classe la première méthode de planification familiale jamais utilisée par un répondant comme étant « moderne » (valeurs 100 à 199) ou « traditionnelle » (valeurs de 200 à 299). Créez une nouvelle variable qui divise le type de méthode en méthodes modernes et traditionnelles en fonction des codes et des fréquences présentées sur le site web de IPUMS-PMA. Recréez l'histogramme ci-dessus deux fois, en vous limitant aux méthodes seulement modernes ou seulement traditionnelles. Spécifiez 35 intervalles pour chacun

